

Appl. No. 09/844,730
Amdt. dated January 21, 2005
Reply to Office action of October 20, 2004

Amendments to the Claims:

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

1. (Currently amended) A method for clustering data ~~in a system having an integrator and at least two computing units comprising:~~

- (a) loading each of a plurality of computing units with common global parameter values and a particular corresponding local data set;
- (b) receiving, by an integrator, from each computing unit, generating local sufficient statistics based on the local data set and global parameter values; and
- (c) employing the local sufficient statistics of all the computing units to update the global parameter values,

wherein the local sufficient statistics and the global parameter values support implementation of a distributed K-Harmonic Means clustering algorithm or a distributed Expectation-Maximization clustering algorithm.

2. (Original) The method of claim 1 wherein the step of loading each computing unit with common global parameter values and a particular local data set further comprises:

- a_1) receiving a set of data points to be clustered;
- a_2) dividing the data points into at least two local data sets;
- a_3) sending common global parameter values to each of the computing units; and
- a_4) sending each local data sets to a designated computing unit,

Appl. No. 09/844,730
Amdt. dated January 21, 2005
Reply to Office action of October 20, 2004

3. (Currently amended) The method of claim 2 wherein the step of employing the local sufficient statistics of all the computing units to update the global parameter values further comprises:

- ~~c_1)~~ ~~each computing unit sending its local sufficient statistics to the integrator;~~
- ~~e_2)~~ the integrator determining global sufficient statistics based on the local sufficient statistics of all the computing units; and
- ~~e_3)~~ c_2) the integrator determining updated global parameter values based on the global sufficient statistics.

4. (Currently amended) The method of claim 1 further comprising:

- d) checking ~~the a~~ convergence quality;
- e) determining whether the convergence quality meets a predetermined quality; and
- f) when the convergence meets a predetermined quality, stop processing; otherwise;
- g) when the convergence fails to meet a predetermined quality, providing the updated global parameter values to the computing units and repeating steps (a) to (c).

5. (Original) The method of claim 2 wherein sending common global parameter values to each of the computing units includes the step of:

broadcasting common global parameter values to each of the computing units.

6. (Original) The method of claim 2 further comprising the step of:

initializing the common global parameter values before sending the common global parameter values to each of the computing units.

7. (Canceled).

Appl. No. 09/844,730
Amdt. dated January 21, 2005
Reply to Office action of October 20, 2004

8. (Original) The method of claim 1 wherein a distributed K -Harmonic Means clustering algorithm is implemented.

9. (Original) The method of claim 1 wherein a distributed Expectation-Maximization (EM) clustering algorithm is implemented.

10. (Original) The method of claim 1 wherein the data points to be clustered are naturally distributed.

11. (Currently amended) A distributed data clustering system comprising:

- (a) a first computing unit that generates a first set of local sufficient statistics for performing data clustering based on global parameter values and a first local data set that is a subset of data points to be clustered and global parameter values to generate first local sufficient statistics;
- (b) a second computing unit that generates a second set of local sufficient statistics for performing data clustering based on global parameter values and a second local data set that is a subset of the data points to be clustered and global parameter values to generate second local sufficient statistics; and
- (c) an integrator unit that receives for receiving the first and second sets of local sufficient statistics from the first and second computing units, respectively, and that employs for employing the first and second local sufficient statistics to update the global parameter values, wherein the global parameter values include centers, covariance matrices, and mixing probabilities in accordance with an Expectation-Maximization (EM) clustering algorithm.

12.-14. (Canceled).

Appl. No. 09/844,730
Amdt. dated January 21, 2005
Reply to Office action of October 20, 2004

15. (Original) The distributed data clustering system of claim 11 wherein the first and second local data sets include data points that are naturally distributed.

16. (Original) The distributed data clustering system of claim 11 wherein the integrator receives a set of data points to be clustered, divides the data points into at least two local data sets, sends common global parameter values to each of the computing units, and sends each of the local data sets to a designated computing unit.

17. (Currently amended) The distributed data clustering system of claim 11, ~~wherein the integrator receives the local sufficient statistics from the first and second computing units; and~~

wherein the integrator ~~determining~~ unit determines global sufficient statistics based on the local sufficient statistics of the first and second computing units; and

wherein the Integrator unit determines updated global parameter values based on the global sufficient statistics.

18. (Currently amended) The distributed data clustering system of claim 11 wherein the integrator ~~checks the convergence quality, determines whether the a~~ convergence quality meets a predetermined quality, and when the convergence meets a predetermined quality, the integrator stops processing, ~~when-but while~~ the convergence fails to meet a predetermined quality, the integrator provides updated global parameter values to the computing units.

19. (Original) The distributed data clustering system of claim 11 wherein the integrator broadcasts common global parameter values to the first and second computing units.

Appl. No. 09/844,730
Amdt. dated January 21, 2005
Reply to Office action of October 20, 2004

20. (Original) The distributed data clustering system of claim 11 wherein the integrator initializes the common global parameter values before sending the common global parameter values to the first and second computing units.

21. (New) A distributed K Harmonic Means clustering system that comprises:
a plurality of computing units each configured to receive a set of centers, and each further configured to combine the set of centers with a local data set to obtain local sufficient statistics for updating the set of centers in accordance with the K Harmonic Means clustering algorithm;

at least one integrator unit configured to combine the local sufficient statistics from each of the plurality of computing units to obtain global sufficient statistics, and further configured to use the global sufficient statistics to update the set of centers in accordance with the K Harmonic Means clustering algorithm.

22. (New) The system of claim 21, wherein the local sufficient statistics include a dynamically-weighted sum of inverse distances for each center in the set of centers, and a dynamically-weighted sum of data components for each center in the set of centers.